In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner

Published online 17 July 2014

Nucleic Acids Research, 2014, Vol. 42, No. 14 ***9493–9503***
doi: 10.1093/nar/gku617

In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner

Gairik Sachdeva[1,2,3], Abhishek Garg[2], David Godding[4], Jeffrey C. Way[3] and Pamela A. Silver[2,3,*]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA, [2]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA, [3]Wyss Institute for Biologically Inspired Engineering, 3 Blackfan St., Boston, MA 02115, USA and [4]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

Received May 22, 2014; Revised June 23, 2014; Accepted June 25, 2014

ABSTRACT

Co-localization of biochemical processes plays a key role in the directional control of metabolic fluxes toward specific products in cells. Here, we employ *in vivo* scaffolds made of RNA that can bind engineered proteins fused to specific RNA binding domains. This allows proteins to be co-localized on RNA scaffolds inside living *Escherichia coli*. We assembled a library of eight aptamers and corresponding RNA binding domains fused to partial fragments of fluorescent proteins. New scaffold designs could co-localize split green fluorescent protein fragments to produce activity as measured by cell-based fluorescence. The scaffolds consisted of either single bivalent RNAs or RNAs designed to polymerize in one or two dimensions. The new scaffolds were used to increase metabolic output from a two-enzyme pentadecane production pathway that contains a fatty aldehyde intermediate, as well as three and four enzymes in the succinate production pathway. Pentadecane synthesis depended on the geometry of enzymes on the scaffold, as determined through systematic reorientation of the acyl-ACP reductase fusion by rotation via addition of base pairs to its cognate RNA aptamer. Together, these data suggest that intra-cellular scaffolding of enzymatic reactions may enhance the direct channeling of a variety of substrates.

INTRODUCTION

Spatial organization of biochemical reactions is commonly observed in biological systems and allows for complex metabolic control. Natural approaches include compartmentalization, separation of reactions in different cells and macromolecular organizations of enzymes (1–6). These strategies control the direction of metabolic pathway flux (1,2), prevent loss of intermediates to competing cellular reactions (3), and insulate the cell itself from intermediates that may be volatile (4). Polyketide and fatty acid synthases are examples of large multi-enzyme complexes where intermediates are shuttled between enzymes to achieve end products with specific lengths and chemical groups (5). Similarly in bacteria, protein microcompartments for propanediol and ethanolamine utilization restrict the diffusion of toxic intermediates such as propionaldehyde and acetaldehyde respectively (6).

Co-localization of enzymes has been successfully employed in synthetic systems as well (7). Confinement of enzymatic reactions has been demonstrated both *in vitro* (8–10) and *in vivo.* In bacteria, both protein (11,12) and DNA (13) have been used as scaffolds for heterologous enzymatic pathways with several fold increases in end product titers reported. Recently, we showed that RNA scaffolds can be used to modularly organize proteins in *Escherichia coli* (14) and increase the rate of electron transfer between enzymes. Such scaffolds, formed by tiles (14) or repeats (15) of small RNA strands, lead to higher densities of binding sites available for protein docking.

The role of enzyme co-localization in pathways that involve diffusible intermediates continues to be debated (16). While some analyses rely on three dimensional diffusion models to argue that the effectiveness of substrate channeling may be limited to cases where intermediate pools are dilute (17), others have suggested a more general 'molecular crowding' hypothesis (18). The latter view is supported by a recent study, in which the authors demonstrated increased

*To whom correspondence should be addressed. Tel: 617-432-6401; Fax: 617-432-5012; Email: pamela_silver@hms.harvard.edu
Present address:
Abhishek Garg, Novartis Institutes for Biomedical Research, Cambridge, MA 02139, USA.

transfer of hydrogen peroxide between enzymes scaffolded *in vitro* (9). In that study, a discontinuous increase in intermediate channeling was observed for proteins tethered <20 nm apart. The authors proposed a mechanism of metabolite channeling by restricted diffusion on hydration layers across crowded protein surfaces. Metabolite diffusion may be particularly restricted for hydrophobic substrates such as fatty acids and related molecules, which may associate with protein surfaces more strongly than hydrophilic substrates. In this regard, the two-step alkane synthesis pathway described by Schirmer *et al*. (19) is of interest because it involves an uncharged aldehyde intermediate. The two enzymes, acyl-acyl carrier protein (ACP) reductase and aldehyde deformylating oxygenase (20) can be heterologously expressed in *E. coli* to produce pentadecane. Also of interest is the channeling of cofactors and carboxylic acids, since they form intermediates in several industrially relevant metabolic products. Succinate is one such product where pathway engineering has focused on increasing supply of oxaloacetate (21), and reduced nicotinamide adenine dinucleotide (NADH) (22). Co-localization of enzymes such that the fatty aldehyde intermediate, or oxaloacetate and NADH are transferred directly from the surface of one enzyme to the other could enhance production of pentadecane and succinate respectively.

Here, we express synthetic RNA strands comprised of polymerization domains and aptamers in bacterial cells (23). The polymerization domains allow for assembly of RNA strands into a macromolecular structure. We show assembly using fluorescent protein fusions to a library of aptamer-binding domain pairs that includes small RNA binding peptides (24) as well as natural domains like MS2 and PP7. Further, fusions of RNA-binding domains, with enzymes involved a two-step pentadecane pathway, are co-expressed to allow assembly onto the RNA scaffolds. We show that increased production of pentadecane depends on the relative geometric orientation of enzymes. Further, we report higher metabolic output on RNA scaffolds with up to four enzymes co-localized from a multi-step succinate production pathway. Our results suggest that the specific orientation and placement of enzymes on the scaffold enhances pathway flux consistent with restricted substrate diffusion.

MATERIALS AND METHODS

Strain construction

E. coli strain BL21* (DE3), that carries a mutation is the *me131* (RNaseE) gene, was used for all experiments described here. Supplementary Table S3 lists the plasmid systems used for expression of the various components in *E. coli.* RNA and protein expression was based on Novagen (EMD Millipore) Duet vectors: pETDuet1 was used for RNA expression because of its high copy number (>40 copies/cell) while the lower copy pACYCDuet1 (10–15 copies/cell) and pCOLADuet1 (20–40 copies/cell) were used to express proteins. For co-expression, we co-transformed competent BL21* DE3 cells with the three relevant plasmids and selected on plates with a combination of antibiotics. For all metabolic production experiments, triplicates correspond to three independent transformant colonies.

Microscopy and analysis

For green fluorescent protein (GFP) microscopy experiments, cells were grown in LB media and induced with 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD 0.8. Cells were allowed to grow for three more hours at 37°C with shaking. They were then washed with M9 media to reduce the background fluorescence from LB media. Cells were placed on 2% clear agarose pads and analyzed using a Nikon TE-2000 microscope (100×, 1.4 numerical aperture objective, ORCA-ER charge-coupled device camera, FITC channel). The images were analyzed using the software packages ImageJ and MicrobeTracker (25,26). We calculated the average cellular fluorescence using MicrobeTracker and plotted the distribution as histograms using the software DataGraph 3.1.1. 2013. Visual Data Tools, Inc.

Metabolic production culture conditions

For alkane production experiments, *E. coli* cells were grown in M9 media with 3% glucose, added minerals and 0.1% Triton X-1000 as described in Supplementry Information for Schirmer *et al*. (19). Cultures were inoculated at OD 0.8 with 1mM IPTG. Five mililiter cultures were grown in 20-ml glass test tubes at 37°C with shaking. Alkane yields were measured 16 h after induction unless otherwise stated. For succiniate production experiments, *E. coli* cells were grown anaerobically in LB media, with 2% glucose, 2% Bis-Tris, 100 mM formate and induction with 2 mM IPTG. For this, overnight LB cultures, still in exponential phase, were centrifuged and resuspended in the above media at a starting OD of 5.0 and the batch cultures were sparged to create a headspace of 16% CO_2 and balance nitrogen. Succinic acid levels were measured 7 h after anaerobic growth at 37°C.

Measuring production levels

Following induction and incubation, cell cultures were centrifuged, and the supernatant decanted. In a typical pentadecane assay, 3 ml of supernatant vigorously mixed with 1 ml of 1M NaCl and 1 ml pure high-performance liquid chromatography (HPLC) grade ethyl acetate (Sigma Aldrich). After 1 h of shaking, the ethyl acetate extract was separated by centrifugation (10 min at 4000 *g*) and obtaining the top layer. Extracts were analyzed by gas chromatography–mass spectrometry (GC-MS) on an Agilent GC-MS 5975/7890 (Agilent Technologies) with the HP-5MS column (30 m length, 0.50 mm diameter). Following a 1–3 μl split-less injection, the inlet temperature was held at 100°C for 3 min and ramped up to 320°C at 30°C per minute. For quantification, standard curves of pentadecane and hexadecanol were generated, by using the pure compounds (Sigma Aldrich). Succinic acid production assays were similarly performed on decanted supernatant. Roche (r-biopharm) enzymatic succinic acid assay kits (CN: 10176281035) were used to measure production levels using a Victor3V 1420 Multilabel plate reader (Perkin Elmer).

Measuring enzyme expression

For enzyme level quantification, we collected centrifuged cell pellets from cultures that were grown in media condition described for alkane production. Cultures grown for 16 h after induction were used. The cell pellets were lysed with a bacterial protein extraction reagent (B-PER II, Thermo Scientific), boiled for 5 min at 100°C, and a protease inhibitor cocktail (Roche) was immediately added. We used 4–20% Tris-Glycine gels (Novex) to separate the proteins by mass, transferred them to a nitrocellulose membrane (Novex iBlot) and blotted with antibodies. We used Abcam antibodies for endogenous GAPDH (ab85760) and 6xHis Tag (ab1187), and Novagen antibody for Strep Tag II (#71591). The software ImageJ (26) was used to analyze and quantify western blot images.

Purification of RNA-bound proteins

For co-purification of RNA interacting with enzymes, we performed UV crosslinking with 150 mJ/cm^2 of *E. coli* cultures in 6-well plates. This was followed by gentle cell lysis using B-PER I (Thermo Scientific) containing protease inhibitors, RNaseOUT (Life Technologies) and DNase I (Sigma Aldrich). The solution was subjected to 2 freeze thaw cycles (−20/4°C) to further ensure cell lysis. Then the lysate was centrifuged at 12 000 *g* for 10 min and the supernatant was processed through the HisPur Ni-NTA magnetic beads protocol (Thermo Scientific) to purify 6xHis-tagged AAR fusion protein. IgA and IgG magnetic beads (Thermo Scientific) were used as controls to pull down random proteins. The purification was verified by western blots and then subjected to protease K (Thermo Scientific) treatment, to release any co-purified RNA.

Measuring cellular RNA Levels

Intracellular RNA was either collected from cultures grown for 16 h after induction or from protease digestion of the purified UV crosslinked enzymes. RNA directly from cell pellets was extracted using Trizol and a Purelink RNA Mini Kit (Ambion, Life Technologies). RNA from a protease treated purified enzyme extract was extracted through acid-phenol-chloroform extraction and overnight ethanol precipitation at −20°C. From this, cDNA was prepared using SuperScript III^{TM} First-Strand cDNA synthesis system (Life Technologies). We used random hexamers to reverse transcribe the entire cellular RNA pool. Five hundred nanogram of total RNA was elongated at 25°C for 10 min, followed by 50°C for 50 min and termination at 85°C for 5 min. RT-PCR was performed using 7.5 μl of SYBR Green Supermix (Ambion), 4.5 μl of 10× diluted cDNA and 3 μl of 2 nM primer solution. Supplementary Table S7 lists the primers that were used for amplification. The gapA primers served to enable measurement of endogenous cellular gapA mRNA pools, which were used as internal controls. The 2D primers amplified scaffold sequences accurately as tested by a standard curve using 4-fold dilutions of a single test sample. Quantitative polymerase chain reaction (PCR) was performed on an Eppendorf Mastercycler ep realplex thermal cycler (upto 55 cycles) and the qPCR software Biogazelle Plus was used for calculating scaffold RNA concentration relative to gapA mRNA. Quantitative PCR efficiency was estimated to be 1.87/cycle for scaffold RNA and 2.03/cycle for gapA using a dilution series. Hence, relative amplification, *r*, for scaffold/gapA per cycle is $r \sim 0.92$.

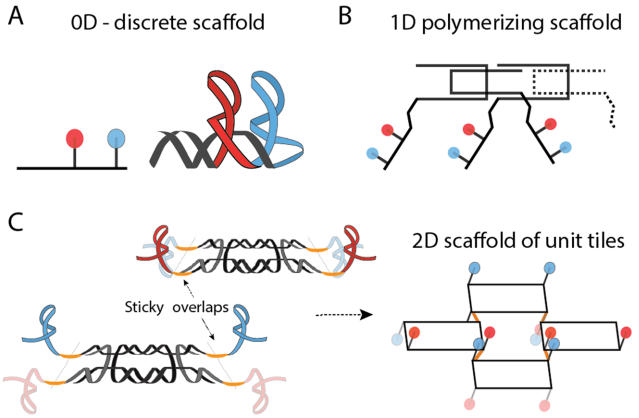


Figure 1. Three dimensionalities of RNA scaffolds. Our designs allow formation of three different types of RNA scaffolds (14). (**A**) In the 0D case, a single RNA strand folds into a discrete unit presenting two aptamers. (**B**) The 1D polymerizing strands have 'sticky' ends that allow the individual units to form linear chains of aptamer sites through complementary base pairing. (**C**) In the case of 2D scaffolds, two different RNA oligonucleotides, A and B, come together to form unit tiles A_2B_2, which then polymerize in two dimensions through sticky end base pairs. Each corner of the unit-tile interactions brings together two different aptamers in close proximity. Red and blue represent different aptamers and dimmed shades show aptamers in a pointing downwards from the scaffold plane.

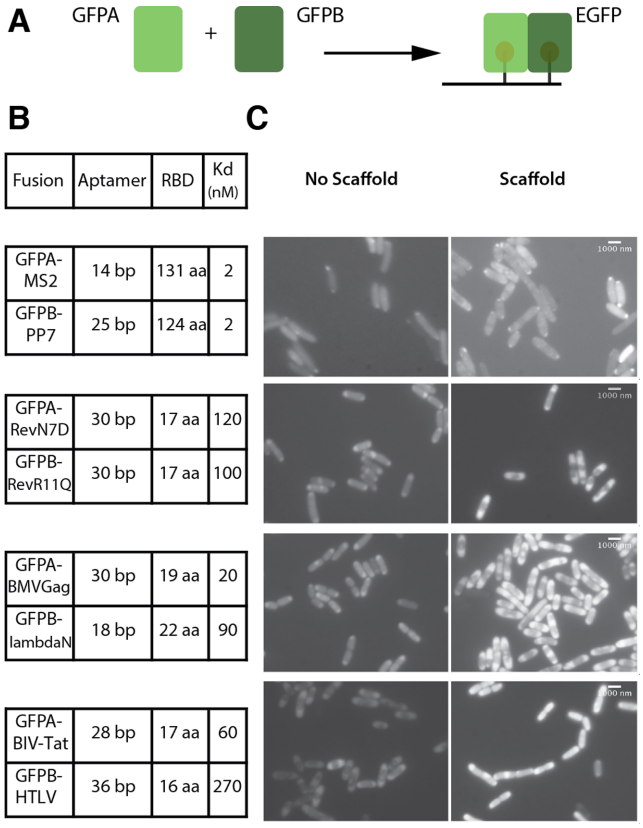
Scaffold structure representation

The Make-NA tool (James Stroud 2004, 2011), available on the University of Southern California servers, was used for constructing PDB representations of our stem and T7 terminator hairpins. The software Autodesk Maya (student version 2014) was used to render the 3D representation, along with a Molecular Maya toolkit (Digizyme).

RESULTS

Expanding the repertoire of functional RNA binding domains

We have expanded the set of RNA binding domains that can be used to form fusion proteins that will bind to specific sites on our RNA scaffolds. The basic strategy builds on a scheme in which engineered *E.coli* express proteins fused to RNA binding domains (RBDs) that would assemble onto RNA scaffolds that include aptamers and polymerization domains. Strand design allows the aptamers to be presented on discrete [0-dimensional (0D)], 1-dimensional (1D) or 2-dimensional (2D) scaffolds (14) (Figure 1). Each protein to be docked binds to its cognate binding site on the RNA scaffold. In previous work, RNA binding domains from viral coat proteins like MS2 and PP7 were fused with fluorescent proteins or enzymes for co-localization on RNA strands (14,15). These viral RBDs bind tightly to their RNA aptamers (27,28) but are relatively large (>120 amino acids) and hence may affect folding or activity of the resulting fusion protein through steric interactions. Bayer



Fusion	Aptamer	RBD	Kd (nM)
GFPA-MS2	14 bp	131 aa	2
GFPB-PP7	25 bp	124 aa	2
GFPA-RevN7D	30 bp	17 aa	120
GFPB-RevR11Q	30 bp	17 aa	100
GFPA-BMVGag	30 bp	19 aa	20
GFPB-lambdaN	18 bp	22 aa	90
GFPA-BIV-Tat	28 bp	17 aa	60
GFPB-HTLV	36 bp	16 aa	270

Figure 2. A library of aptamer-RNA binding domain pairs can be functionally expressed *invivo*. (**A**) Schematic for split GFP complementation on an RNA scaffold (**B**) List of four different pairs—with eight unique aptamer-RNA binding sets tested. Values of dissociation constants reported previously (24,27,28) are tabulated with domain and aptamer sizes. (**C**) FITC images showing enhanced split GFP fluorescence *invivo* in the presence of RNA scaffolds for all sets tested. Scale bars: 1 μm.

et al. (24) characterized several short arginine-rich motifs that bind RNA along with their synthetically selected RNA aptamers. We chose six of these short RBD–aptamer sets, along with MS2 and PP7, to create a library of eight such interactions for *in vivo* applications (Figure 2B). The six sets were picked so that we could obtain a set of binding domains that were both short (<25 amino acids), to enable facile tagging on enzymes, and also mutually orthogonal in binding to the corresponding aptamers (24) (Supplementary Tables S1 and S2).

To test *in vivo* RNA–protein interactions for the library obtained above, we made fusions of the RBDs in four pairs with inactive green fluorescent protein fragments (split GFP) A and B respectively (Figure 2). The fragments were split as described by Valencia-Burton *et al*. (29). Corresponding aptamers for each pair of RBDs were included in RNA sequences, with their secondary structures designed to fold into discrete scaffolds (Figure 2A). These '0D' scaffolds were co-expressed with split GFP–RBD fusions through plasmids in *E. coli*. In control cells, we expressed only the split GFP–RBD fusions with an empty vector so that no scaffolds for co-localizing the fragments were present. After 3 h of induction, the cells were visualized at 100× magnification under a microscope to assay cellular fluorescence indicating co-localization of the split GFPs to create active GFP.

Protein–RNA binding *in vivo* was in evidence for all eight aptamer–RBD sets tested in four pairs (Figure 2C). When the split GFP fragments are expressed without scaffolds (Figure 2C—left column) the cells are only dimly fluorescent. Upon co-expression with the scaffold sequence, however, significantly brighter cells are observed, presumably because the split GFP fragments bind adjacent to each other on the RNA scaffold (Figure 2C—right column). Histograms showing average cellular fluorescein isothiocyanate (FITC) intensity from 100 to 400 cells for the four pairs are presented in Supplementary Figure S3. We have previously shown such split GFP complementation on RNA scaffolds for the PP7 and MS2 domains' pair (14), which served as a positive control here. We also observed enhanced GFP complementation on scaffolds from three pairs of the six new aptamer–RBD partners. This gives us a tool-kit of eight potential aptamer–protein interaction sets with varying sizes and binding properties.

RNA binding domains function on polymerizing RNA scaffolds

We showed that the split GFP complementation is efficient on RNA scaffolds of higher dimensions (1D and 2D) with the new RBDs and aptamers as well. For this we chose two aptamer–RBD interaction sets: PP7 and BIV-Tat. We chose PP7 because of its tight binding and ability to form non-aggregating dimers on RNA aptamers (28). BIV-Tat was chosen because it was the strongest-interacting short domain for which a crystal structure of the aptamer–RBD complex is known (24,30).

As above, we expressed fused split GFP fragments with a plasmid expressing either no RNA or RNA strands for 0D, 1D or 2D RNA scaffolds (Figure 3A). Fluorescence microscope images of the cells were analyzed using software packages MicrobeTracker (25) and ImageJ (26) to quantify average fluorescence per unit area for several images (>100 cells) of each strain. We performed statistical analyses (Figure 3B) to calculate the average cellular fluorescence intensity in each case. A progressive shift toward brighter cells upon expression of RNA scaffolds of the three dimensionalities was observed (Figure 3A,B). The automated analysis indicated two apparent populations when the GFP fragments are co-expressed with the 2D scaffold; visual inspection of the fields confirmed that there are two populations of bright cells. We hypothesize that the kinetics of RNA tiling lead to extended 2D scaffold assembly in the brightest population of cells either due to small differences in RNA expression levels or stochastic coalescing of multiple assemblies.

Increased pentadecane production on RNA scaffolds

To use RNA scaffolds for metabolic substrate channeling, we employed a two-step pentadecane production pathway (Figure 4A) which has recently been characterized using cyanobacterial enzymes expressed in *E. coli* (19). Two enzymes, acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO) (20) [formerly known as aldehyde decarbonylase (DCB) (19)] sequentially convert cellular N-

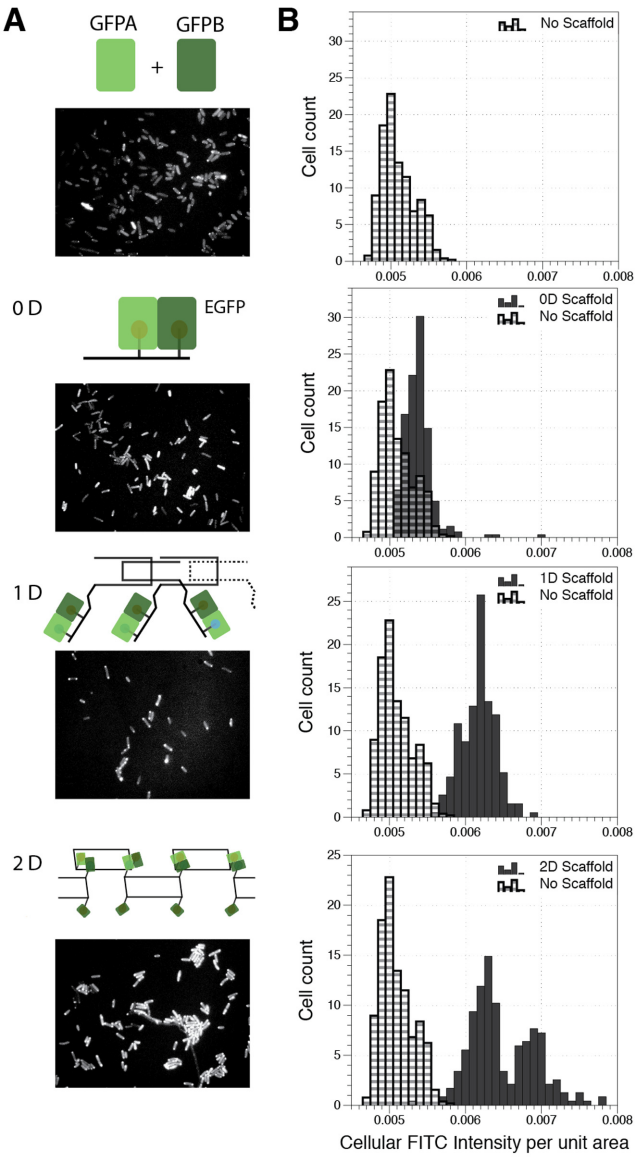


Figure 3. Aptamers incorporated into 0D, 1D and 2D RNA scaffolds bind proteins *invivo* (**A**) FITC images showing GFP fluorescence observed in cells on expression of split GFP fragments (GFPA-PP7 and GFPB-BIV-TAT) with either no scaffold, 0D, 1D or 2D scaffolds containing PP7 and BIV-TAT binding aptamers (**B**) Corresponding FITC quantification (histograms) for 0, 1 and 2D scaffold designs ($P < 10^{-3}$, one-tailed *t*-test).

carbon fatty acyl-ACPs to alkanes of length N-1. For example, AAR reduces the C16 acyl-ACP molecule to produce a fatty aldehyde, hexadecanal ($C_{15}H_{31}CHO$), which undergoes deformylation catalyzed by ADO to form pentadecane. However, upon co-expression of these two enzymes in *E. coli*, the major products are alcohols and not alkanes (19). The intermediate aldehyde appears to be be non-specifically reduced by aldehyde reductases in the cell (31). We hypothesized that if hexadecanal could be selectively channeled from AAR to ADO on scaffolds, an increase in pentadecane would occur.

Genetic fusions of AAR and ADO to RNA binding domains were constructed on the N-termini of the enzymes. AAR was tagged with BIV-Tat and ADO was fused to PP7; both fusions retained activity (Figure 4C), although attempted fusions with the MS2 failed to show activity. *E. coli* expressing the two enzyme fusions were grown and alkane levels in the supernatant were measured ('Materials and Methods' section). Pentadecane production, which is absent in wild-type *E. coli*, was observed (Figure 4C)—confirming catalytic activity of the fusion proteins. We measured alkane production after 16 h of log-phase growth in batch cultures, while earlier work has reported higher levels can be achieved from up to 40 h of induced growth (19). In addition, we also observed other alkane products however, levels were much smaller than pentadecane, consistent with previous reports (19), and could not be accurately quantified.

When the enzyme fusions were co-expressed with 1D, or 2D RNA scaffolds containing their cognate binding sites (Figure 4B) in *E. coli*, alkane production was enhanced by ~80% compared to co-expression with an empty vector or a 0D scaffold, as assayed by pentadecane levels in ethyl acetate extracts (Figure 4C). We also measured production of hexadecanol, one of the alcohol side products reported when AAR and ADO are co-expressed (19). We observe that the hexadecanol levels are reduced when enzymes are co-localized on their cognate 2D scaffolds, compared to expression without any RNA scaffolds (Figure 4D). As a control, expression of the enzymes with a 2D scaffold lacking the cognate aptamers showed no difference from the empty vector control (Figure 4C and D). The mismatched scaffold carried binding sites for RevR11Q and MS2 instead of the BIV-TAT and PP7 aptamers. Lack of any effect on the mismatched scaffold shows that co-expression with the correct cognate scaffold was necessary for an effect.

We also sought direct evidence of enzyme–RNA scaffold interaction in our co-expression system. For this, we UV-crosslinked protein–RNA interactions in cells expressing the pentadecane pathway enzymes and co-purified BIV-TAT-AAR with any bound RNA (Figure 4E). BIV-TAT-AAR was tagged with 6xHistidine and could be seen only upon purification with Ni-NTA magnetic beads and not when random IgA + IgG beads were used as control (Figure 4E). After digestion of the protein and RNA purification, we looked for presence of scaffold RNA levels by quantitative RT-PCR. Internal primers were designed to amplify scaffold RNA and as well cellular gapA mRNA. While scaffold RNA and gapA are both detected in unpurified cell lysates expressing 2D scaffolds, only the scaffold RNA is detected from the extracts co-purified with BIV-TAT-AAR. Neither gapA nor scaffold RNA are detected from the IgA + IgG pull downs. This suggests that while a control mRNA like gapA does not interact with our enzymes appreciably, interactions of BIV-TAT-AAR and the co-expressed 2D scaffold are significantly higher.

To ensure that the enhanced synthesis reflected in Figure 4C was due to scaffolding of the enzymes and not differential expression, we compared protein levels using Strep and 6xHis tags included in the fusions for AAR and ADO respectively. Quantitative western blots were prepared on cell extracts and enzyme levels were found to be unaffected by co-expression with scaffolds (Figure 4F). This indicates that when we expressed the two-enzyme pathway in the presence

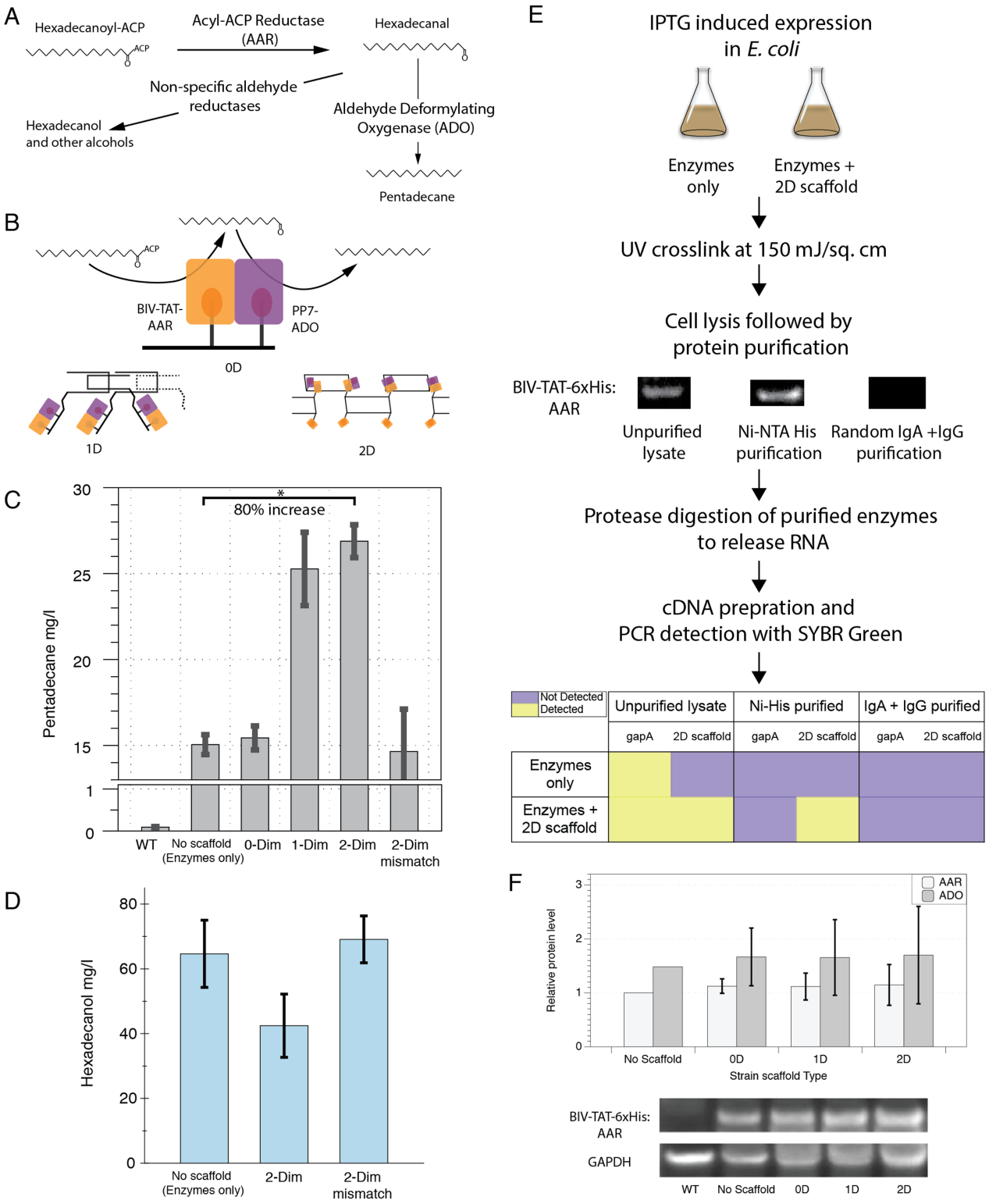


Figure 4. Localizing enzymes on *in vivo* RNA scaffolds leads to increased metabolic flux (**A**) The alkane synthesis pathway with the aldehyde reductase side reaction (19). (**B**) Scaffolding designs implemented to channel the hexadecanal intermediate toward alkane production. (**C**) Synthesis of alkanes from *E. coli* expressing PP7-ADO, BIV-Tat-AAR, an empty vector or RNA scaffolds of zero, one or two dimensions (with aptamers for BIV-TAT and PP7), and a 2D scaffold with mismatched aptamers (containing anti-RevR11Q and MS2). (**D**) One of the pentadecane pathway side-products, hexadecanol, production measured on expression of pathway enzymes with the same empty vector, correct 2D scaffold and mismatched 2D scaffold. (**E**) Direct evidence of enzyme–RNA interaction from experimental flow for detecting RNA associated *invivo* with 6xHis tagged BIV-TAT-AAR enzymes. Differently shaded boxes indicate whether amplification was detected by SYBR green or RNA species were below detection threshold. (**F**) Pathway enzyme production levels measured by quantitative western blotting using 6xHis (AAR) and Strep (ADO) tags and normalized to GAPDH. Levels are shown for co-expression with an empty vector or scaffolds of different dimensionalities ($n = 3$, error bars = SEM) (* indicates $P < 0.05$, one-tailed t-test).

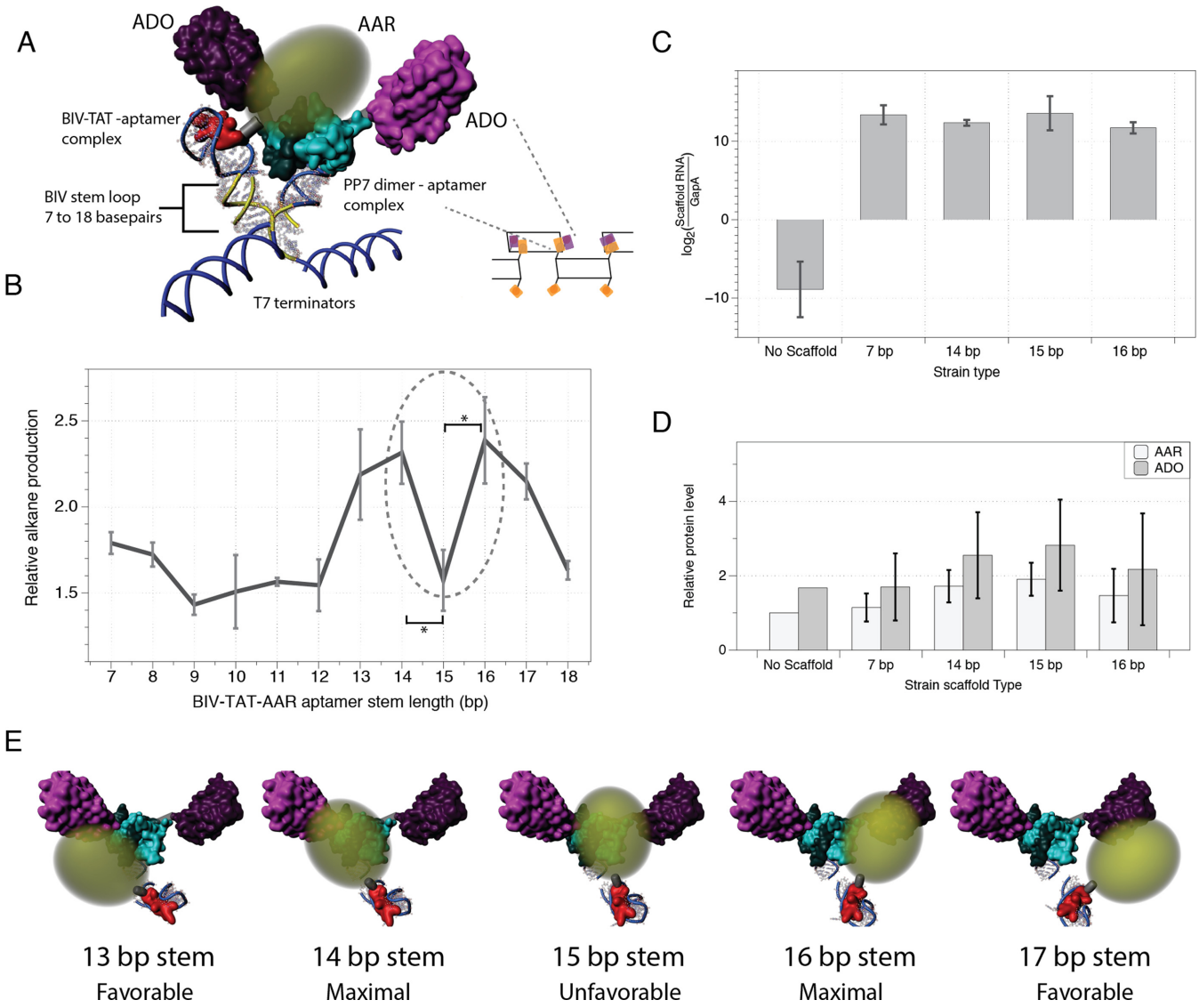


Figure 5. Length and orientation of aptamers affect metabolic flux increase on scaffolds. (**A**) Schematic showing design details of a pair of aptamers interacting on the 2D scaffold using PDB IDs: 2QUX (34), 2A9X (35) and 4KVQ (36). Different base pair lengths (7–18bp) were used for the hairpin presenting aptamer bound by BIV-Tat-AAR so that rotation of the complex may bring AAR in closer proximity to ADO on the scaffold. (**B**) Pentadecane yields are shown for 2D scaffolds with the different BIV-Tat binding aptamer stem lengths. Concentrations were normalized to yield from enzyme expression without scaffolds and plotted as relative production levels. (**C**) RNA scaffold production levels, measured by qRT-PCR and normalized to endogenous gapA mRNA, for the strains showing greatest variation in yields. (Log base value $2r \sim 1.84$—see 'Materials and Methods' section). (**D**) Pathway enzyme production levels measured by quantitative western blotting using 6xHis (AAR) and Strep (ADO) tags. Levels are shown for co-expression of pathway enzymes with RNA scaffolds of varying BIV-TAT-aptamer lengths and are normalized to GAPDH levels. (**E**) Proposed model for two maximal configurations of intermediate flux channeling. On varying the anti-BIV-TAT aptamer stem loop length, different rotational conformations of the BIV-Tat-AAR moiety are possible, relative to the PP7-ADO dimer ($n = 3$, error bars = SEM) (* indicates $P < 0.05$, one-tailed t-test).

of 1D or 2D polymerizing RNA scaffolds, higher pentadecane titers were realized for the same amounts of proteins expressed.

Flux enhancement depends on geometry of scaffolded enzymes

The RNA scaffold offers structural flexibility that might be further modified to enhance the interplay of associated enzymes. Some studies have probed the effect of varying numbers of binding sites (11) or different inter-enzyme distances on metabolic channeling (9). Here, we investigate the effect of altering relative enzyme orientation on the scaffold.

The structure of 2D sheets (Figure 1) resulting from assembly of the RNA oligonucleotides (Supplementary Table S6) that make up the scaffold can be inferred based on the helicity of double-stranded RNA (11 base pair (bp)/turn), the structure of Holliday junctions (32), and general principles of nucleic acid nanotechnology (14,33). Four strands (A_2B_2) come together to form a unit tile, which polymerizes in two dimensions through base pair interactions on the tile corners. The aptamers are presented on two different sides

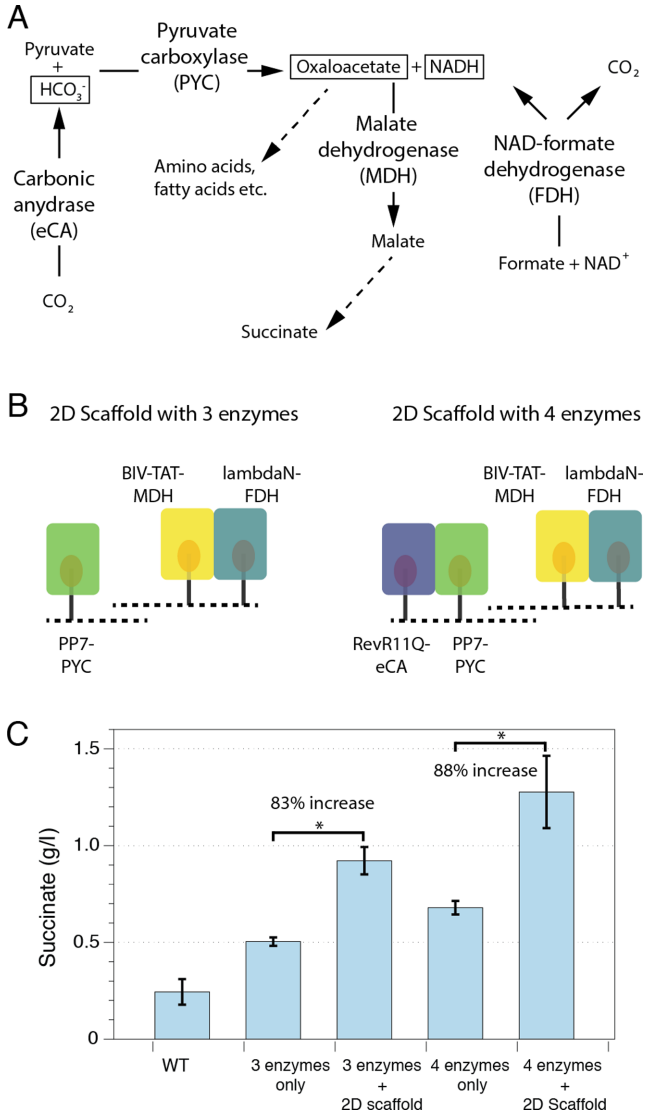


Figure 6. Enhanced succinate production on scaffolds. (**A**) Enzymes expressed to increase succinate production in *E. coli* with the intermediate molecules shown in boxes. (**B**) Schematics for two scaffolding approaches on 2D scaffolds to increase pathway flux toward succinate. PYC and FDH are scaffolded with MDH in the first approach to prevent loss of oxaloacetate and NADH respectively. In the scaffold with four aptamers, eCA is also scaffolded to provide HCO_3^- for PYC. (**C**) Succinate production from expressing three enzyme fusions (PP7-PYC, BIV-TAT-MDH, lambaN-FDH) or four enzymes (also RevR11Q-eCA) with and without the corresponding 2D scaffolds compared to wild-type (WT) levels ($n = 3$, error bars = SEM) (* indicates $P < 0.05$, one-tailed t-test).

of the plane of the unit tile, as shown (Figures 1 and 5A). Each unit tile has four 6-bp sticky ends that base pair with diagonally related units.

Through these interactions, one PP7 aptamer and one BIV-Tat aptamer emerge on the same side of the scaffold plane at each corner of the unit tiles (Figure 5A; Supplementary Information 2.2). We constructed a high-level model of this aptamer pair from PDB files of aptamer-binding protein complexes, double-stranded RNAs, and the ADO enzyme from a related organism (34–36). The enzymes being co-localized on the scaffold bind and form PP7-aptamer (PDB ID: 2QUX (34)) and BIV-Tat–aptamer (PDB ID: 2A9X (35)) complexes. Such complexes can be treated as rigid elements composed of PP7-linker-ADO and BIV-Tat-linker-AAR fusion proteins which are expected to be under constrained Brownian motion as predicted by the coarse-grained molecular dynamics simulation approach of Robinson-Mosher *et al.* (37). In Figure 5A, the crystal structure of ADO is represented by *Prochlorococcus marinus* aldehyde-deformylating oxygenase (PDB ID: 4KVQ (36)) which has 95% sequence identity to the *Synechococcus elongatus* enzyme used here.

We found that systematic addition of base pairs to the stem-loop structure of anti-BIV-Tat aptamer identified scaffold variants with increased activity. To explore the effect of changing the position of the reductase, we added 0–11 bp to the BIV-Tat hairpin on the 2D scaffold, to go through an entire RNA helical turn (Figure 5A). We then measured the pentadecane synthesis from cells expressing the 122D scaffolds with the BIV-Tat-AAR and PP7-ADO fusion proteins (Figure 5B). As before, the strains were grown aerobically, induced, and finally extracted with ethyl acetate for assaying pentadecane production. The enzymes co-expressed with an empty vector served as control. All designs showed increased pentadecane production relative to the no-scaffold case (Figure 5B). The increase in production ranges from ~40% (9 bp) to ~140% (16 bp) greater than co-expression without scaffolds.

A striking pattern of pentadecane synthesis was observed as a function of hairpin stem length, particularly with stems of 13–17 bp. Synthesis was near-maximal with 13- and 17-bp stems, maximal with 14- and 16-bp stems, and minimal with a 15-bp stem (Figure 5B). These results may be interpreted in terms of a model (Figure 5E; 'Discussion' section) where the peaks in activity with stems of lengths 13, 14, 16 and 17 bp may correspond to the acyl-ACP reductase being in proximity to one or the other of the deformylating oxygenases fused to the dimeric PP7 protein.

To ensure that these effects were solely due to changes in scaffold design, we measured the relative levels of scaffold RNA and enzyme production for strains with the 7, 14, 15 and 16 bp scaffolds. We extracted total cellular RNA from cell pellets for each strain after an induction experiment and measured scaffold RNA levels by quantitative RT-PCR. Once again, primers to measure the 2D scaffold RNA level and endogenous cellular pools of gapA mRNA were used (Figure 5C). All measured scaffold RNAs were produced at about the same levels, with a variation of only 2 to 4-fold, corresponding to the error of qPCR. Enzyme levels were measured by quantitative western blots on cell pellet extracts from these strains (Figure 5D); no significant difference in enzyme concentrations between the strains and the no-scaffold controls were seen.

Localizing multiple enzymes on RNA scaffolds

RNA scaffolds can be used to enhance pathway flux in longer multi-step enzymatic pathways. We chose the commercially relevant succinate production pathway that has multiple rate limiting control points at the pyruvate utilization node (38) (Figure 6A). Production of oxaloacetate through pyruvate carboxylase (PYC) enhances flux toward

succinate (21). Further, increased succinate production is noted when a NAD^+-dependent formate dehydrogenase (FDH) generates NADH, a co-factor used in the conversion of oxaloacetate to malate, by malate dehydrogenase (MDH) (22). Thus, we wanted to test if scaffolding of PYC (from *Rhizobium etli)*, MDH (from *E. coli*), and FDH (from *Candida boidinii*) on a 2D scaffold results in increased redirection of pyruvate toward succinate in *E. coli* (Figure 6B). We also tested a four-enzyme co-localization scheme, in which a cyanobacterial carbonic anhydrase (eCA), known to increase oxaloacetate flux toward succinate (39) was included (Figure 6B). We made fusions of PYC, MDH, FDH and eCA with RNA binding domains PP7, BIV-TAT, lambdaN and RevR11Q respectively. We observe that co-expression of PP7-PYC, BIV-TAT-MDH and lambdaN-FDH with a 2D scaffold containing the corresponding aptamers leads to 83% higher succinate production than expression of the enzyme fusions with an empty vector (Figure 6C). When RevR11Q-eCA is also expressed with these three enzymes, the new corresponding scaffold with four aptamers leads to an 88% increase (Figure 6C). These results indicate that intermediates as diverse in their chemical composition as oxaloacetate, NADH and HCO_3^- can likely be channeled toward desired product formation on RNA scaffolds with 2–4 different modular aptamer sites.

DISCUSSION

Compartmentalization and scaffolding of enzymes are effective strategies for the control of cellular metabolism in both natural (1–6) and synthetic (1,7–14) systems. In this work we established a toolkit for designing protein–RNA interactions that work inside cells and used it to enhance metabolic substrate channeling in a biofuel production pathway. The key features of our system are the expression in cells of RNAs that form a multidimensional nanostructure scaffold with binding sites for RNA-binding domains (aptamers) and co-expression of fusions between specific RBDs and enzymes defining a metabolic pathway.

We show *in vivo* functionality of a library of modular RNA binding domains and aptamer scaffolds. These RBD–aptamer sets are effective in co-localizing proteins within living *E. coli.* The RBD–aptamer sets were used to reconstitute split GFP fragments on a variety of RNA scaffolds. We showed assembly of 0D, 1D and 2D RNA scaffolds, which were progressively more effective in sub-cellular co-localization of proteins. We then used these tools to demonstrate metabolic channeling on RNA scaffolds inside cells. A two-enzyme pentadecane synthesis pathway with a labile aldehyde intermediate was localized on the scaffolds. Scaffolds with 2D assemblies significantly increased pathway yields. We showed direct evidence of enzyme–RNA interaction and systematically reoriented one of the enzymes on the scaffold. We found that, in the best case, a 2D RNA scaffold led to 2.4-fold higher pentadecane production levels than observed without scaffolding. We could also demonstrate the utility of RNA scaffolds to enhance flux through 3–4 steps of a succinate production pathway with diverse intermediates.

The library of eight RNA binding domain-aptamer sets tested here forms a useful toolkit for designing *in vivo* protein–RNA interactions. The library of protein/RNA pairs is orthogonal and includes a range of aptamer lengths, binding domain sizes and binding dissociation constants. Characterization of shorter RNA binding peptides will, in principle, enable facile engineering of enzyme fusions with minimal perturbation to protein structure. We have also demonstrated the modular nature of our RNA scaffold design, by using the aptamer site library and showing assembly of functional discrete, 1D or 2D scaffolds.

Our system offers a convenient, modular design to optimize metabolic channeling by changing one design element at a time. In Figure 4C, pentadecane yield is enhanced much more on 1D and 2D scaffolds than the 0D design. This result could be due to non-optimal relative orientations of enzymes, and insufficient metabolic channeling on the 0D scaffold. In the 2D scaffold, the enzymes AAR and ADO are associated with aptamers emerging on the same side of the unit tile plane, where they are well positioned to interact with each other. At each corner junction between unit tiles, 6-bp sticky ends bring together two different stem-loop aptamer segments, followed by RNA tails corresponding to the T7 terminator sequence at the 3′ end of the transcripts (Supplementary Figure S4). Because six bases correspond to half of a turn of an RNA helix, these are likely positioned on the same side of the plane of the scaffold, and electrostatic repulsion between the aptamers, the T7 terminator RNA, and the scaffold itself will roughly determine the placement of the two interacting aptamers. In the solved structure of the PP7 protein–RNA complex (PDB ID: 2QUX (34)) the two PP7-ADO units are at an angle of ∼120°. The BIV-Tat–aptamer interaction is a one-to-one interaction (PDB ID: 1MNB (30)). Our data suggest that for some conformations of the BIV-Tat-AAR moiety that are separated by about 120°, the AAR interacts favorably with the one or the other of the PP7-ADO dimers, while other rotational conformations of the aptamer make transfer of the metabolic intermediate less favorable (Figure 5E). Since the atomic structure of AAR has not been solved, we are unable to consider active site orientations, which might be able to help explain our observations with higher precision.

The geometric dependence of RNA scaffold performance that we report here points to a mechanism where phenomena other than just 3D spatial diffusion of metabolites (17) are in play. Our results are consistent with direct, surface-limited metabolite transfer between aligned enzyme active sites (9). The range of flux enhancement on 2DRNA scaffolds reported here illustrates the importance of getting the precise relative orientation right when co-localizing enzymes. In the future, mathematical modeling of protein surface diffusion phenomena will shed insights into how our results and those of other groups (9,18) can be mechanistically explained. Experimentally investigating the behavior of substrates with different sizes and hydrophobicity could also help develop such a model.

The economics of biofuel synthesis are particularly challenging because biofuels compete with petroleum products that derive from carbon fixed over millions of years (40). Our results indicate that rational scaffolding of enzymes can be useful in improving yields of pathways with intermediates of a variety of different chemical properties. The 2.4-fold increase in pentadecane synthesis observed here could

be synergistically combined with other genetic techniques for biofuel yield enhancement (41), leading to higher overall synthesis rates.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online including [1–11].

ACKNOWLEDGMENT

The authors thank the following for support during preparation of this article: Qingqing Wang, Tyler J. Ford, Jessica Polka, Joe P. Torella and Buz Barstow for helpful discussions during the project; and Krista Shapton for contributions to illustrations.

ACCESSION NUMBERS

PDB IDs: 2QUX, 2A9X, 4KVQ and 1MNB.

FUNDING

Harvard University Center for Environment and Link Foundation Energy Fellowship (to G.S.); Swiss National Science Foundation Fellowship (to A.G.); DARPA Living Foundries Program [4500000572 to P.A.S.]. Funding for open access charge: DARPA Living Foundries Program [4500000572 to P.A.S.].
Conflict of interest statement. None declared.

REFERENCES

1. Agapakis,C.M., Boyle,P.M. and Silver,P.A. (2012) Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.*, **8**, 527–535.
2. Yeates,T.O., Kerfeld,C.A., Heinhorst,S., Cannon,G.C. and Shively,J.M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.*, **6**, 681–691.
3. Srere,P.A. (1987) Complexes of sequential metabolic enzymes. *Annu. Rev. Biochem.*, **56**, 89–124.
4. Penrod,J.T. and Roth,J.R. (2006) Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. *J. Bacteriol.*, **188**, 2865–2874.
5. Smith,S. and Tsai,S.-C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.*, **24**, 1041–1072.
6. Yeates,T.O., Crowley,C.S. and Tanaka,S. (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.*, **39**, 185–205.
7. Chen,A.H. and Silver,P.A. (2012) Designing biological compartmentalization. *Trends Cell Biol.*, **22**, 662–670.
8. Erkelenz,M., Kuo,C.-H. and Niemeyer,C.M. (2011) DNA-mediated assembly of cytochrome P450 BM3 subdomains. *J. Am. Chem. Soc.*, **133**, 16111–16118.
9. Fu,J., Liu,M., Liu,Y., Woodbury,N.W. and Yan,H. (2012) Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *J. Am. Chem. Soc.*, **134**, 5516–5519.
10. You,M., Wang,R.-W., Zhang,X., Chen,Y., Wang,K., Peng,L. and Tan,W. (2011) Photon-regulated DNA-enzymatic nanostructures by molecular assembly. *ACS Nano*, **5**, 10090–10095.
11. Dueber,J.E., Wu,G.C., Malmirchegini,G.R., Moon,T.S., Petzold,C.J., Ullal,A.V., Prather,K.L.J. and Keasling,J.D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.*, **27**, 753–759.
12. Moon,T.S., Dueber,J.E., Shiue,E., Prather,K.L.J. and Prather,K.L. (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.*, **12**, 298–305.
13. Conrado,R.J., Wu,G.C., Boock,J.T., Xu,H., Chen,S.Y., Lebar,T., Turnsek,J., Tomsic,N., Avbelj,M., Gaber,R. *et al.* (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.*, **40**, 1879–1889.
14. Delebecque,C.J., Lindner,A.B., Silver,P.A. and Aldaye,F.A. (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science*, **333**, 470–474.
15. Golding,I. and Cox,E.C. (2004) RNA dynamics in live Escherichia coli cells. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 11310–11315.
16. Idan,O. and Hess,H. (2013) The origins of activity enhancement in enzyme cascades on scaffolds. *ACS Nano*, **7**, 8658–8665.
17. Barros,L.F. and Martínez,C. (2007) An enquiry into metabolite domains. *Biophys. J.*, **92**, 3878–3884.
18. Lee,H., DeLoache,W.C. and Dueber,J.E. (2012) Spatial organization of enzymes for metabolic engineering. *Metab. Eng.*, **14**, 242–251.
19. Schirmer,A., Rude,M.A., Li,X., Popova,E. and del Cardayre,S.B. (2010) Microbial biosynthesis of alkanes. *Science*, **329**, 559–562.
20. Warui,D.M., Li,N., Nørgaard,H., Krebs,C., Bollinger,J.M. Jr and Booker,S.J. (2011) Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *J. Am. Chem. Soc.*, **133**, 3316–3319.
21. Vemuri,G.N., Eiteman,M.A. and Altman,E. (2002) Effects of growth mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of Escherichia coli. *Appl. Environ. Microbiol.*, **68**, 1715–1727.
22. Berrios-Rivera,S.J. *et al.* (2002) Metabolic engineering of Escherichia coli: increase of NADH availability by overexpressing an NAD+ -dependent formate dehydrogenase. *Metab. Eng.*, **4**, 217–229.
23. Delebecque,C.J., Silver,P.A. and Lindner,A.B. (2012) Designing and using RNA scaffolds to assemble proteins in vivo. *Nat. Protoc.*, **7**, 1797–1807.
24. Bayer,T.S., Booth,L.N., Knudsen,S.M. and Ellington,A.D. (2005) Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*, **11**, 1848–1857.
25. Sliusarenko,O., Heinritz,J., Emonet,T. and Jacobs-Wagner,C. (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.*, **80**, 612–627.
26. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
27. Spingola,M. and Peabody,D.S. (1997) MS2 coat protein mutants which bind Qbeta RNA. *Nucleic Acids Res.*, **25**, 2808–2815.
28. Lim,F.F., Downey,T.P.T. and Peabody,D.S.D. (2001) Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. *J. Biol. Chem.*, **276**, 22507–22513.
29. Valencia-Burton,M., McCullough,R.M., Cantor,C.R. and Broude,N.E. (2007) RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat. Methods*, **282**, 296–298.
30. Puglisi,J.D., Chen,L., Blanchard,S. and Frankel,A.D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR Peptide-RNA complex. *Science*, **270**, 1200–1203.
31. Reiser,S. and Somerville,C. (1997) Isolation of mutants of Acinetobacter calcoaceticus deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *J. Bacteriol.*, **179**, 2969–2975.
32. Eichman,B.F., Vargason,J.M., Mooers,B.H.M. and Ho,P.S. (2000) The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 3971–3976.
33. Aldaye,F.A. and Sleiman,H.F. (2009) Supramolecular DNA nanotechnology. *Pure Appl. Chem.*, **81**, 2157–2181.
34. Chao,J.A., Patskovsky,Y., Almo,S.C. and Singer,R.H. (2008) Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.*, **15**, 103–105.
35. Leeper,T.C., Athanassiou,Z., Dias,R.L.A., Robinson,J.A. and Varani,G. (2005) TAR RNA recognition by a cyclic peptidomimetic of Tat Protein. *Biochemistry*, **44**, 12362–12372.
36. Khara,B., Menon,N., Levy,C., Mansell,D., Das,D., Marsh,E.N.G., Leys,D. and Scrutton,N.S. (2013) Production of propane and other short-chain alkanes by structure-based engineering of ligand

specificity in aldehyde-deformylating oxygenase. *ChemBioChem*, **14**, 1204–1208.
37. Robinson-Mosher,A., Shinar,T., Silver,P.A. and Way,J. (2013) Dynamics simulations for engineering macromolecular interactions. *Chaos*, **23**, 025110-1–025110-13
38. Sánchez,A.M., Bennett,G.N. and San,K.-Y. (2005) Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. *Metab. Eng.*, **7**, 229–239.
39. Wang,D., Li,Q., Li,W., Xing,J. and Su,Z. (2009) Improvement of succinate production by overexpression of a cyanobacterial carbonic anhydrase in Escherichia coli. *Enzyme Microb. Technol.*, **45**, 491–497.
40. Sims,R.E.H., Mabee,W., Saddler,J.N. and Taylor,M. (2010) An overview of second generation biofuel technologies.*Bioresour. Technol.*, **101**, 1570–1580.
41. Gronenberg,L.S., Marcheschi,R.J. and Liao,J.C. (2013) Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.*, **17**, 1–10.